Property-driven Machine Learning

Thomas Flinkow
Supervisors: Rosemary Monahan and Barak A. Pearlmutter

June 2025

In recent years, a range of formal verification tools has emerged to ensure that neural networks adhere to logical specifications—a key requirement for their deployment in safety-critical domains.

Despite these advances, neural networks have been shown to frequently fail to meet specifications after training. Perhaps not surprisingly, we cannot expect neural networks to infer logical properties purely from data.

A promising approach is *property-driven machine learning*, which incorporates logical specifications directly into training by expressing them as additional loss terms. Many such loss functions—known as *differentiable logics*—have been proposed, including those based on classical multi-valued logic systems like fuzzy logic.

Our work so far has investigated differentiable logics with respect to their suitability for machine learning tasks. We recently developed a software framework to support property-driven machine learning in practice. Future work is to move beyond empirically improving adherence to logical specifications and toward methods that offer formal guarantees.